PROCEEDINGS *of the* FIFTH BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY

Held at the Statistical Laboratory
University of California
June 21–July 18, 1965
and
December 27, 1965–January 7, 1966

with the support of
University of California
National Science Foundation
National Institutes of Health
Air Force Office of Scientific Research
Army Research Office
Office of Naval Research

VOLUME IV

BIOLOGY AND PROBLEMS OF HEALTH

EDITED BY LUCIEN M. LE CAM
AND JERZY NEYMAN

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY AND LOS ANGELES
1967

THE DISTRIBUTION OF THE TOTAL SIZE OF AN EPIDEMIC

H. E. DANIELS
UNIVERSITY OF BIRMINGHAM

1. Introduction

This paper examines in some detail the distribution of the total number of cases in an epidemic of the general stochastic type for a closed population. The assumed model is that of Bartlett [2] and McKendrick [11] which Bailey [1] used to study the stochastic analogue of the deterministic threshold theorem (Kermack and McKendrick [10], D. G. Kendall [9]). Bailey obtained recurrence relations from which the required probabilities were computed numerically. His calculations revealed a gradual transition from J-shaped distributions containing only small epidemics for population sizes below the threshold, to U-shaped distributions containing either large or small epidemics but practically no epidemics of intermediate size when the threshold is exceeded. There is also an interesting transitional form of distribution near the threshold value.

In an attempt to understand what motivates an epidemic to behave in this way, Whittle [13] and Kendall [9] constructed different models approximating to the one used by Bailey but easier to handle analytically. Both explained Bailey's results in terms of an initial birth and death process where extinction is certain in the first case and not certain in the second. This work is summarized, with additional references, in the book by Bailey [2]. In a paper presented at this Symposium, Gani [7] develops some recent work by Siskind [12] and himself [6] on a method of obtaining time dependent solutions of the epidemic equations. For the limiting case considered here he shows how the probabilities can be computed by successive multiplication of matrices.

My main object is to arrive at approximate formulae for the distribution of total epidemic size which are appropriate for large populations. The approach differs from that of most other investigations in that the backward equations of the process are used. (See, however, Bartlett [3].) I also find it convenient to work in terms of the number remaining uninfected, rather than the total number of new cases. The technique by which the approximations are obtained was previously used by me in an entirely different context (Daniels [4], [5]). As presented here it should not be regarded as rigorously establishing the approximations, though I have no doubt that the results are correct and numerical comparisons bear this out.

2. The deterministic model

Suppose that at time t there are x susceptibles, y infectives and z recovered or dead in the population. Initially we shall take $x = \xi$, $y = \eta$, $z = 0$, so that $x + y + z = \xi + \eta$. The deterministic epidemic has been fully studied by Kendall [9], but we examine it briefly for the sake of some results referred to later. The deterministic equations are, with a suitable time scale,

$$\frac{dx}{dt} = -xy, \qquad \frac{dy}{dt} = xy - \rho y, \qquad \frac{dz}{dt} = \rho y, \tag{2.1}$$

where ρ is called the *relative removal rate* by Bailey, and the *threshold* by Kendall. Then $dx/dz = -x/\rho$ and $x = \xi \exp(-z/\rho)$ for all t. At the end of the epidemic $y = 0$, $z = \xi + \eta - x$, and the number x of individuals remaining uninfected satisfies the equation

$$x \exp(-x/\rho) = \xi \exp[-(\xi + \eta)/\rho]. \tag{2.2}$$

We suppose that ξ and ρ are large and η/ρ is small. There are two values of x satisfying (2.2) near the respective roots ξ, ξ' of $x \exp(-x/\rho) = \xi \exp(-\xi/\rho)$. The only relevant root is the one less than ξ. When $\xi < \rho$, this root is near ξ and (2.2) gives as a first approximation

$$x = \xi - \eta\xi/(\rho - \xi). \tag{2.3}$$

On the other hand, when $\xi > \rho$ the required root is near $\xi' < \rho$ and approximately,

$$x = \xi' - \eta\xi'/(\rho - \xi'). \tag{2.4}$$

These hold provided ξ is not near ρ, though it is easy to get an approximate transitional form for $\xi \sim \rho$. Together, they constitute the deterministic threshold theorem in terms of the number remaining uninfected. Notice that when $\xi \gg \rho$ then ξ'/ρ is small and

$$x \sim \xi' \sim \xi e^{-\xi/\rho}. \tag{2.5}$$

3. The stochastic model

In the continuous time model considered by Bartlett and McKendrick the transitions in $(t, t + \delta t)$ are $(x, y) \rightarrow (x - 1, y + 1)$ with probability $xy\delta t + o(\delta t)$ and $(x, y) \rightarrow (x, y - 1)$ with probability $\rho y \delta t + o(\delta t)$. As we are concerned only with the final distribution of x, it is simpler to work with the random walk of transitions in the x, y plane such that

$$\begin{aligned} P\{(x, y) \rightarrow (x - 1, y + 1)\} &= x/(\rho + x), \\ P\{(x, y) \rightarrow (x, y - 1)\} &= \rho/(\rho + x), \end{aligned} \tag{3.1}$$

where initially $x = \xi$, $y = \eta$ and absorption occurs on $y = 0$.

An alternative formulation of the random walk in terms of the numbers of new cases $w = \xi - x$ and removals z is of some interest. This is

$$(3.2)\qquad \begin{aligned} P\{(w, z) \to (w + 1, z)\} &= (\xi - w)/(\rho + \xi - w), \\ P\{(w, z) \to (w, z + 1)\} &= \rho/(\rho + \xi - w). \end{aligned}$$

The problem can then be described in terms of a game involving a mixture of sampling with and without replacement. A box contains ξ real pennies and ρ false ones. The player starts with a capital of η pennies and the price of a draw is one penny. If he draws a false penny he replaces it in the box. If he draws a real penny he keeps it and is allowed a further trial free. The game stops when the player is ruined $(w + \eta = z)$ or when he has drawn all the real pennies $(w = \xi)$.

Let $p(x|\xi, \eta)$ be the probability that there are ultimately x uninfected individuals, when initially there were ξ susceptibles and η infectives. The backward equations for p are

$$(3.3)\qquad \xi p(x|\xi - 1, \eta + 1) + \rho p(x|\xi, \eta - 1) - (\rho + \xi)p(x|\xi, \eta) = 0,$$

$\xi > x$, $\eta \geqq 1$, and

$$(3.4)\qquad \rho p(x|x, \eta - 1) - (\rho + x)p(x|x, \eta) = 0$$

with the condition

$$(3.5)\qquad p(x|\xi, 0) = \delta(\xi - x),$$

where $\delta(\xi - x) = 0$, $\xi \neq x$ and $\delta(0) = 1$. From (3.4),

$$(3.6)\qquad p(x|x, \eta) = [\rho/(\rho + x)]p(x|x, \eta - 1) = \cdots = [\rho/(\rho + x)]^{\eta}.$$

Our method of attack depends on the fact that

$$(3.7)\qquad \binom{\xi}{x + s}\left(\frac{\rho}{\rho + x + s}\right)^{\xi - x + \eta}$$

is a solution of (3.3) for arbitrary s, and satisfies (3.6) when $s = 0$. We try to build up a solution of the form

$$(3.8)\qquad p(x|\xi, \eta) = \sum_{s=0}^{\xi - x} A_s \binom{\xi}{x + s}\left(\frac{\rho}{\rho + x + s}\right)^{\xi - x + \eta}$$

satisfying the required conditions. If $A_0 = 1$, condition (3.6) is satisfied, and from (3.5) we must have

$$(3.9)\qquad \delta(\xi - x) = \sum_{s=0}^{\xi - x} A_s \binom{\xi}{x + s}\left(\frac{\rho}{\rho + x + s}\right)^{\xi - x}.$$

The coefficients A_s can be determined recursively and hence $p(x|\xi, \eta)$ can be found, provided A_s is independent of ξ. That this is so becomes clear if we write

$$(3.10)\qquad A_s = (-)^s\binom{x + s}{s}H_s,$$

and use the fact that

$$\binom{\xi}{x+s}\binom{x+s}{\xi} = \binom{\xi}{x}\binom{\xi-x}{s}. \tag{3.11}$$

Then,

$$p(x|\xi, \eta) = \binom{\xi}{x}\sum_{s}^{\xi-x} (-)^s H_s \binom{\xi-x}{s}\left(\frac{\rho}{\rho+x+s}\right)^{\xi-x+\eta}, \tag{3.12}$$

$$\delta(\xi - x) = \sum_{s=0}^{\xi-x} (-)^s H_s \binom{\xi-x}{s}\left(\frac{\rho}{\rho+x+s}\right)^{\xi-x}. \tag{3.13}$$

The coefficients H_s depend only on x and ρ.

As a method of computing the probabilities, these equations are if anything less convenient than the original equations (3.3), (3.4), or the corresponding forward equations, when ξ is large and $\xi > \rho$. Their value lies in the fact that a technique is available for obtaining an asymptotic approximation when ξ is large.

4. Some exact results

The problem is substantially simplified by using the following result. Since the left side of (3.13) is zero except when $\xi = x$, it can equally well be put in the form

$$\delta(\xi - x) = \sum_{s=0}^{\xi-x} (-)^s H_s \binom{\xi-x}{s}\left(\frac{\rho+x}{\rho+x+s}\right)^{\xi-x}. \tag{4.1}$$

If we then write $H_s \equiv H_s(x, \rho)$ and $p(x|\xi, \eta) \equiv p(x|\xi, \eta, \rho)$ to show their dependence on ρ, it appears that

$$H_s(x, \rho) = H_s(0, \rho + x), \tag{4.2}$$

and from (3.13) we get the exact relation

$$p(x|\xi, \eta, \rho) = \left(\frac{\rho}{\rho+x}\right)^{\xi-x+\eta}\binom{\xi}{x} p(0|\xi - x, \eta, \rho + x). \tag{4.3}$$

(H. F. Downton has succeeded in deriving (4.3) by a direct combinatorial argument.) This enables everything we want to be deduced from a knowledge only of the behavior of $p(0|\xi, \eta, \rho)$ which we now study. The equations are, with H_s for $H_s(0, \rho)$,

$$p(0|\xi, \eta, \rho) = \sum_{s=0}^{\xi} (-)^s H_s \binom{\xi}{s}\left(\frac{\rho}{\rho+s}\right)^{\xi+\eta}, \tag{4.4}$$

$$\delta(\xi) = \sum_{s=0}^{\xi} (-)^s H_s \binom{\xi}{s}\left(\frac{\rho}{\rho+s}\right)^{\xi}. \tag{4.5}$$

Taking $\xi = 0, 1, 2, \cdots$ in (4.5) we get a set of equations for H_s which can be solved in determinantal form as

$$(4.6)\qquad H_s = \frac{s!}{c_s^s}\begin{vmatrix} c_0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \frac{c_0^2}{2!} & \frac{c_1}{1!} & 1 & 0 & 0 & & 0 & 0 \\ \frac{c_0^3}{3!} & \frac{c_1^2}{2!} & c_2 & 1 & 0 & & 0 & 0 \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & & \cdot \\ \frac{c_0^{s-1}}{(s-1)!} & \frac{c_1^{s-2}}{(s-2)!} & \frac{c_2^{s-3}}{(s-3)!} & & & & c_{s-2} & 1 \\ \frac{c_0^s}{s!} & \frac{c_1^{s-1}}{(s-1)!} & \frac{c_2^{s-2}}{(s-2)!} & \cdot & \cdot & \cdots & \frac{c_{s-2}^2}{2!} & c_{s-1} \end{vmatrix},$$

where $c_s = \rho/(\rho + s)$ and $H_0 = 1$. Substitution in (4.4) gives an explicit solution for $p(0|\xi, \eta, \rho)$.

5. Approximations below the threshold

The form of the solution (4.6) bears an unexpected and, I suspect, fortuitous resemblance to expressions for the probabilities associated with the Poisson process with a curved absorbing boundary and the related Kolmogorov-Smirnov test [5]. For example, it follows from the work on that problem that H_s can also be expressed in the form $H_s = c_s^{-s}\, C_s(0)$, where

$$(5.1)\qquad C_s(z) = s! \int_z^{c_0} dz_1 \int_{z_1}^{c_1} dz_2 \cdots \int_{z_{s-1}}^{c_{s-1}} dz_s$$

is a so called Gontcharoff polynomial [8] whose jth derivative vanishes at $z = c_j$. A good deal can be discovered about the asymptotic behavior of these expressions by using a technique originally developed for a related problem [4]. But there is an essential difference here which complicates matters. In the applications mentioned c_s is an *increasing* positive sequence and this ensures that $C_s(0)$ is always positive. In the present problem c_s is a *decreasing* positive sequence and beyond a certain value of s, H_s begins to oscillate with increasing amplitude and alternating sign. Nevertheless, we shall use a modified version of the same technique to study the asymptotic form of H_s and, hence, deduce that of $p(0|\xi, \eta, \rho)$ when ξ and ρ are large. The presentation is heuristic and to some extent incomplete. A rigorous development must depend on a more extensive study of the asymptotic properties of Gontcharoff polynomials which it is hoped to publish later.

In (4.5) replace ξ by m, multiply it by $(-\lambda)^m \binom{\xi}{m}$ and sum from $m = 0$ to ξ. After a little manipulation the result is

$$(5.2)\qquad 1 = \sum_{s=0}^{\xi} H_s \binom{\xi}{s} (\lambda c_s)^s (1 - \lambda c_s)^{\xi - s}, \qquad c_s = \rho/(\rho + s),$$

where λ is an arbitrary parameter. The technique is to look at the behavior of

$$T_s(\lambda) = \binom{\xi}{s} (\lambda c_s)^s (1 - \lambda c_s)^{\xi - s} \tag{5.3}$$

for large ξ. It can be shown to have limiting forms analogous to the normal and Poisson limits of the binomial distribution, with a peak at the unique root s_0 of $s = \lambda \xi c_s$. By varying λ we can scan H_s with this "window" and deduce its asymptotic behavior. The range $0 < \lambda < 1$ ensures that $T_s(\lambda)$ is positive, but λ can be allowed to exceed unity provided $T_s(\lambda)$ remains positive near the root s_0.

We shall consider only the normal limiting form. Assume ξ and ρ are large and write $z = s/\xi$, $dz = 1/\xi$, $c(z) = c_s$, $T_s(\lambda) = T(z, \lambda) dz$. If neither z nor $1 - z$ is small, Stirling's approximation leads to

$$\begin{aligned} T(z, \lambda) &\sim \left[\frac{\xi}{2\pi z(1 - z)}\right]^{1/2} \left[\frac{\lambda c(z)}{z}\right]^{\xi z} \left[\frac{1 - \lambda c(z)}{1 - z}\right]^{\xi(1-z)} \\ &= \left[\frac{\xi}{2\pi z(1 - z)}\right]^{1/2} \left\{1 - \frac{[z - \lambda c(z)]}{z}\right\}^{\xi z} \left\{1 + \frac{[z - \lambda c(z)]}{1 - z}\right\}^{\xi(1-z)} \\ &= \left[\frac{\xi}{2\pi z(1 - z)}\right]^{1/2} \exp\left\{-\frac{\xi[z - \lambda c(z)]^2}{2z(1 - z)} + \begin{array}{l}\text{terms involving}\\ \text{higher powers}\\ \text{of } z - \lambda c(z)\end{array}\right\}. \end{aligned} \tag{5.4}$$

The maximum of $T(z, \lambda)$ is at the unique root z_0 of $z - \lambda c(z) = 0$, and it is shown in [4] that there are no other maxima. Near z_0 we can write $z - \lambda c(z) \doteqdot (z - z_0)[1 - \lambda c'(z_0)]$, and because $z - z_0$ is $O(\xi^{1/2})$ over its effective range, we get the normal approximation

$$T(z, \lambda) \sim \left[\frac{\xi}{2\pi z_0(1 - z_0)}\right]^{1/2} \exp\left\{-\frac{\xi[1 - \lambda c'(z_0)]^2}{2z_0(1 - z_0)} (z - z_0)^2\right\}, \tag{5.5}$$

where

$$z_0 = \lambda c(z_0). \tag{5.6}$$

If it can now be assumed that $H(z) = H_{\xi z}$ varies slowly with z near z_0, then (5.2) approximates to

$$\begin{aligned} 1 &\sim \int_0^1 H(z)\, T(z, \lambda) dz \sim H(z_0) \int_0^1 T(z, \lambda) dz \\ &\sim H(z_0)/[1 - \lambda c'(z_0)] = H(z_0)/[1 - z_0 c'(z_0)/c(z_0)]. \end{aligned} \tag{5.7}$$

Hence,

$$H(z) \sim 1 - zc'(z)/c(z), \tag{5.8}$$

provided z is a possible root of (5.6) at which (5.5) holds. In terms of s, we have $c'(z)/c(z) = -\xi/(\rho + s)$, and we arrive at the approximation

$$H_s \sim 1 + s/(\rho + s) = 2 - \rho/(\rho + s). \tag{5.9}$$

The result is, of course, suggested rather than established by this kind of

reasoning. It depends on the assumption that $H(z)$ varies slowly, and we would expect it to fail for large values of s corresponding to roots for which λ causes $T_s(\lambda)$ to oscillate. Also, small values of s have been excluded by the argument. Nevertheless, calculations show that the approximation is good for values of s up to about ρ (see table I). It should therefore provide an approximation for $p(x|\xi, \eta, \rho)$ when ξ is less than the threshold ρ.

TABLE I

EXACT AND APPROXIMATE VALUES H_s FOR $\rho = 20$
VARIATION OF K_s WITH s

s	H_s	$1 + s/(\rho + s)$	K_s
0	1.0000	1.0000	1.0000
1	1.0500	1.0476	−21.0000
2	1.0948	1.0909	109.4762
3	1.1352	1.1304	−336.3698
4	1.1721	1.1667	732.5702
5	1.2059	1.2000	−1234.8024
6	1.2369	1.2308	1696.7042
7	1.2655	1.2593	−1966.9936
8	1.2921	1.2857	1971.5763
9	1.3168	1.3103	−1740.1758
10	1.3397	1.3333	1371.8941
11	1.3612	1.3548	−977.0948
12	1.3813	1.3750	634.5692
13	1.4002	1.3939	−378.7168
14	1.4179	1.4118	209.0655
15	1.4347	1.4286	−107.3579
16	1.4504	1.4444	51.5285
17	1.4655	1.4595	−23.2195
18	1.4792	1.4737	9.8550
19	1.4938	1.4872	−3.9585
20	1.5028	1.5000	1.5028
21	1.5268	1.5122	−0.5480
22	1.4979	1.5238	0.1840
23	1.6522	1.5349	−0.0664
24	1.1344	1.5455	0.0143
25	3.1684	1.5556	−0.0120
26	−4.8683	1.5652	−0.0053
27	28.2780	1.5745	−0.0086
28	−112.890	1.5833	−0.0091
29	508.531	1.5918	−0.0106
30	−2314.03	1.6000	−0.0121
31	10901.2	1.6079	−0.0137
32	−52810.9	1.6154	−0.0155
33	263176	1.6226	−0.0175
34	−1347561	1.6296	−0.0197
35	7084284	1.6364	−0.0221

Let us substitute (5.8) in the right side of (4.5) to see how nearly it is satisfied for $\xi > 0$. (It is exact for $\xi = 0$). We have

$$\sum_{s=0}^{\xi} (-)^s H_s \binom{\xi}{s} \left(\frac{\rho}{\rho+s}\right)^{\xi} \sim \sum_{s=0}^{\xi} (-)^s \binom{\xi}{s} \left[2\left(\frac{\rho}{\rho+s}\right)^{\xi} - \left(\frac{\rho}{\rho+s}\right)^{\xi+1}\right] \tag{5.10}$$

$$= \frac{1}{(\xi-1)!}\int_0^\infty (2u^{\xi-1} - u^{\xi}/\xi)$$
$$(1 - e^{-u/\rho})^{\xi} e^{-u} du$$
$$= \frac{1}{(\xi-1)!}\int_0^\infty e^{-u}[2u^{2\xi-1}/\rho^{\xi}$$
$$- (1/\xi + \xi/\rho)u^{2\xi}/\rho^{\xi} + \cdots] du.$$

The expansion within the integrand is convergent for small enough u, and termwise integration yields an asymptotic expansion in powers of ρ^{-1} for fixed ξ. The term in $\rho^{-\xi}$ vanishes and the leading term is

$$-\xi(2\xi)!/\rho^{\xi+1}(\xi-1)!. \tag{5.11}$$

For large ρ this is small even at $\xi = 1$. As ξ increases it becomes approximately

$$-[\rho e^2/8(2)^{1/2}](4\xi/\rho e)^{\xi+2} \tag{5.12}$$

which decreases to a minimum at about $\xi = \rho/4$. It does not become appreciable again until ξ approaches $\rho e/4$ after which it rapidly becomes large. So at least for $\xi < \rho e/4$ one can with some confidence insert (5.9) into (4.4) and obtain in the same way,

$$p(0|\xi, \eta, \rho) \sim \frac{1}{(\xi+\eta-1)!}\int_0^\infty [2u^{\xi+\eta-1} - u^{\xi+\eta}/(\xi+\eta)] \tag{5.13}$$
$$(1 - e^{-u/\rho})^{\xi} e^{-u} du$$
$$\sim \frac{\eta(2\xi+\eta-1)!}{(\xi+\eta)!\,\rho^{\xi}} + O(\rho^{-\xi-1}).$$

Then from (4.3) we get for values of ξ below the threshold ρ,

$$p(x|\xi, \eta, \rho) \sim \frac{\eta\rho^{\xi-x+\eta}}{(\rho+x)^{2\xi-2x+\eta}}\binom{\xi}{x}\frac{(2\xi-2x+\eta-1)!}{(\xi-x+\eta)!} \tag{5.14}$$
$$= \frac{\eta\rho^{w+\eta}}{(\rho+\xi-w)^{2w+\eta}}\binom{\xi}{w}\frac{(2w+\eta-1)!}{(w+\eta)!}$$

in terms of the number w of new cases. Since ξ is large, a further approximation leads to the result

$$p \sim \frac{\eta\rho^{w+\eta}\xi^{w}}{(\rho+\xi)^{2w+\eta}}\frac{(2w+\eta-1)!}{w!(w+\eta)!}, \tag{5.15}$$

which is the solution for the birth and death process proposed by Bartlett and exploited by Whittle and Kendall, having birth rate ξ, death rate ρ, and whose mean is given by the deterministic approximation (2.3).

6. Approximations above the threshold

To investigate the range of ξ above the threshold ρ, it is necessary to study the behavior of H_s at values of s beyond those which can be reached by the previous method. It has been mentioned that H_s begins to oscillate violently when s exceeds a certain value. This suggests that the following transformation of (5.2) will be useful for such values of s. Write $\lambda = -\nu/(1-\nu)$, and

$$H_s = (-)^s[(1-c_s)^s/c_s^s]K_s = (-)^s(s/\rho)^s K_s. \tag{6.1}$$

Then, (5.2) becomes

$$(1-\nu)^\xi = \sum_{s=0}^{\xi} K_s \binom{\xi}{s} [\nu(1-c_s)]^s[1-\nu(1-c_s)]^{\xi-s}. \tag{6.2}$$

We could now try using the previous technique with $T_s(\lambda)$ replaced by

$$U_s(\nu) = \binom{\xi}{s} [\nu(1-c_s)]^s[1-\nu(1-c_s)]^{\xi-s}, \tag{6.3}$$

which has peaks at the roots of

$$s = \nu\xi(1-c_s) = \nu s\xi/(\rho+s). \tag{6.4}$$

The lower root $s \doteq 0$ is irrelevant because we are only interested in using (6.2) for large values of s. The upper root is $s_0 = \nu\xi - \rho$, and by varying ν over a suitable range we could examine the behavior of K_s, provided it can be assumed to vary slowly (see table I), after substituting the known approximation (5.9) in (6.1) and (6.2) to cover the lower range of s.

We shall adopt an approach which is based on this idea but is rather more direct for the present problem. Let

$$H_s = 1 + s/(\rho+s) + (-)^s(s/\rho)^s L_s. \tag{6.5}$$

We have seen that the effect of L_s can be ignored at least for $s < \rho e/4$. On substituting (6.5) in (4.5), we get

$$\begin{aligned}\delta(\xi - x) &\sim \sum_{s=0}^{\xi} \left(1 + \frac{s}{\rho+s}\right)(-)^s \binom{\xi}{s} \left(\frac{\rho}{\rho+s}\right)^\xi \\ &\quad + \sum_{s=0}^{\xi} L_s \binom{\xi}{s} \left(\frac{s}{\rho+s}\right)^s \left(\frac{\rho}{\rho+s}\right)^{\xi-s} \\ &= A + B,\end{aligned} \tag{6.6}$$

say. The second term B can be expressed as $\sum L_s U_s(1)$. As before, put $z = s/\xi$ and $U_s(1) = U(z, 1)dz$. The upper root of (6.4) is $s_0 = \xi - \rho$ and provided this is far from zero, $U(z, 1)$ will have an isolated peak at $z_0 = 1 - \rho/\xi$ near which

$$U(z, 1) \sim [\xi/2\pi z_0(1-z_0)]^{1/2} \exp\{-\xi[1 + c'(z_0)]^2(z-z_0)^2/2z_0(1-z_0)\}. \tag{6.7}$$

Assuming that $L_s = L(z)$ varies slowly near z_0, we can infer that

$$B \sim L(z_0)/[1 + c'(z_0)] = \rho L_{\xi-\rho}/(\xi - \rho). \tag{6.8}$$

Consider first the case where ξ is much greater than ρ which is itself large. Under these conditions there is a remarkably simple approximation for $p(x|\xi, \eta, \rho)$ in the region of large epidemics. The terms in A die away rapidly and their sum approximates to

$$A \sim \sum_{s=0}^{\infty} (-\xi e^{-\xi/\rho})^s/s! = \exp(-\xi e^{-\xi/\rho}). \tag{6.9}$$

Hence,

$$\rho L_{\xi-\rho}/(\xi - \rho) \sim -\exp(-\xi e^{-\xi/\rho}). \tag{6.10}$$

The formulae for $p(0|\xi, \eta, \rho)$ corresponding to (6.6) is

$$\begin{aligned} p(0|\xi, \eta, \rho) \sim & \sum_{s=0}^{\xi} \left(1 + \frac{s}{\rho + s}\right)(-)^s \binom{\xi}{s} \left(\frac{\rho}{\rho + s}\right)^{\xi+\eta} \\ & + \sum_{s=0}^{\xi} L_s \binom{\xi}{s} \left(\frac{s}{\rho + s}\right)^s \left(\frac{\rho}{\rho + s}\right)^{\xi-s+\eta}. \end{aligned} \tag{6.11}$$

Suppose that η is small. Compared with (6.6), to the order of approximation considered the effect of the extra factor $[\rho/(\rho + s)]^\eta$ is to leave the first term unaltered and to multiply the second term by $[\rho/(\rho + s_0)]^\eta = (\rho/\xi)^\eta$. It follows that

$$p(0|\xi, \eta, \rho) \sim [1 - (\rho/\xi)^\eta] \exp(-\xi e^{-\xi/\rho}). \tag{6.12}$$

Then from (4.3) we obtain, for small values of x,

$$p(x|\xi, \eta, \rho) \sim [1 - (\rho/\xi)^\eta] \frac{(\xi e^{-\xi/\rho})^x}{x!} \exp(-\xi e^{-\xi/\rho}). \tag{6.13}$$

In other words, when the threshold is large but the population size is much larger, *the distribution of the number remaining uninfected in a large epidemic has approximately the Poisson form with the deterministic mean* $\xi e^{-\xi/\rho}$.

This could perhaps have been conjectured from (4.3) on the plausible supposition that $p(0|\xi - x, \eta, \rho + x)$ changes slowly with x when x is small and ξ is far above the threshold ρ. It is a good approximation for large values of ξ/ρ, but otherwise it is rather a crude fit. One feels that there must be a direct argument in terms of the epidemic process itself to explain this Poissonlike behavior, just as the approximating birth and death process explains the behavior for small epidemics.

At the other end of the distribution near $x = \xi$, (5.14) or (5.15) provides a good approximation. Notice that (5.15) can also be expressed as

$$p \sim (\rho/\xi)^\eta \frac{\eta \rho^w \xi^{w+\eta}}{(\rho + \xi)^{2w+\eta}} \frac{(2w + \eta - 1)!}{w!(w + \eta)!}, \tag{6.14}$$

in accordance with the point of view adopted by Kendall and Whittle, $(\rho/\xi)^\eta$ being the probability that a large epidemic will not develop.

7. A more refined approximation

When ξ is not much larger than ρ the argument of the previous section breaks down principally because the peak of $U(z, 1)$ at $1 - \rho/\xi$ is no longer sufficiently isolated from the one at $z = 0$, and the overlap begins to matter near the tail of the distribution when $\xi - x$ approaches $\rho + x$. Provided this does not happen, we can still improve the previous approximation considerably by looking again at the exact expression for the first term of (6.6) which is, from (5.10),

$$(7.1) \qquad A = \frac{1}{(\xi - 1)!}\int_0^\infty u^{\xi-1}(2 - u/\xi)(1 - e^{-u/\rho})^\xi e^{-u}\, du$$

$$= \frac{\xi^\xi}{(\xi - 1)!}\int_0^\infty v^{\xi-1}(2 - v)(1 - e^{-v\xi/\rho})^\xi e^{-\xi v}\, dv.$$

TABLE II

DISTRIBUTION OF THE FINAL NUMBER x OF UNINFECTED
p is exact; p_1 is from (7.4); p_2 is from (6.13);
only the lower end of the distribution is tabulated.

	$\xi = 1000$	$\eta = 1$	$\rho = 200$	$\xi = 100$	$\eta = 1$	$\rho = 25$
	Exact	Best Fit	Poisson	Exact	Best Fit	Poisson
x	p	p_1	p_2	p	p_1	p_2
0	.00144	.00144	.00095	.13363	.13341	.12012
1	.00821	.00820	.00639	.19039	.18991	.22001
2	.02416	.02413	.02152	.16379	.16321	.20148
3	.04889	.04882	.04834	.11187	.11134	.12303
4	.07646	.07636	.08142	.06741	.06700	.05633
5	.09857	.09843	.10972	.03781	.03752	.02063
6	.10906	.10890	.12322	.02036	.02017	.00630
7	.10649	.10632	.11860	.01073	.01061	.00165
8	.09364	.09349	.09989	.00561	.00553	.00031
9	.07532	.07520	.07479	.00293	.00288	.00008
10	.05609	.05599	.05039	.00154	.00151	.00001
11	.03905	.03898	.03087	.00082	.00080	—
12	.02562	.02557	.01733	.00044	.00043	
13	.01595	.01592	.00898	.00024	.00023	
14	.00947	.00945	.00432	.00013	.00013	
15	.00539	.00538	.00194	.00008	.00007	
16	.00296	.00295	.00082	.00003	.00004	
17	.00157	.00156	.00032	.00002	.00002	
18	.00081	.00080	.00012	.00001	.00001	
19	.00040	.00040	.00004	.00001	.00001	
20	.00020	.00020	.00001	—	.00001	
21	.00009	.00009	.00001		—	
22	.00004	.00004	—			
23	.00002	.00002	—			
24	.00001	.00001	—			

Applying the Laplace method of approximation to the integral, we get

$$A \sim \frac{(2 - v_0)v_0^{\xi}e^{\xi(1-v_0)}(1 - e^{-v_0\xi/\rho})^{\xi}}{\{1 + (v_0 - 1)[(1 + \xi/\rho)v_0 - 1]\}^{1/2}} = f(v_0, \xi, \rho), \tag{7.2}$$

where

$$e^{v_0\xi/\rho} - 1 = (v_0\xi/\rho)/(v_0 - 1). \tag{7.3}$$

(When ξ/ρ is large, $v_0 \sim 1$ and this reduces to (6.9)). Since η is small, the effect on A of the extra factor $[\rho/(\rho + s)]^{\eta}$ in (6.11) is to multiply A by v_0^{η} to the same order of approximation. By the previous argument, we then get for large epidemics

$$p(x|\xi, \eta, \rho) \sim \left(\frac{\rho}{\rho + x}\right)^{\xi - x + \eta} \binom{\xi}{x} \left[v_0^{\eta} - \left(\frac{\rho + x}{\xi - x}\right)^{\eta}\right] f(v_0, \xi - x, \rho + x), \tag{7.4}$$

where v_0 satisfies (7.3) with $\xi - x$ for ξ and $\rho + x$ for ρ. Table II shows (7.4) to be a remarkably good fit even for values as low as $\rho = 25$, $\xi = 100$. When $\rho = 50$, $\xi = 100$ the fit is still found to be good for small values of x but it begins to deteriorate in the tail of the distribution because of the overlap effect mentioned. There seems to be no simple way of allowing for this with the present technique.

From the practical point of view, agreement to such a high order of accuracy is not particularly important because the underlying model is itself very much idealized, and exact computations can in any case be carried out by computer on the original backward or forward equations. But it does give one considerable confidence in the method used to arrive at the approximations.

I am much indebted to Mr. R. L. Holder and Dr. V. D. Barnett for calculations on our English Electric KDF.9. computer which have sustained me throughout the work, and in particular for the computations of tables I and II.

REFERENCES

[1] N. T. J. Bailey, "The total size of a general stochastic epidemic," *Biometrika*, Vol. 40 (1953), pp. 177–185.

[2] ———, *The Mathematical Theory of Epidemics*, London, Griffin, 1957.

[3] M. S. Bartlett, "Some evolutionary stochastic processes," *J. Roy. Statist. Soc. Ser. B*, Vol. 11 (1949), pp. 211–229.

[4] ———, *Stochastic Population Models*, London, Methuen, 1960.

[5] H. E. Daniels, "The statistical theory of the strength of bundles of threads," *Proc. Roy. Soc. Ser. A*, Vol. 183 (1945), pp. 405–435.

[6] ———, "The Poisson process with a curved absorbing boundary," *Bull. Inst. Internat. Statist.*, 34th session, Vol. 40 (1963), pp. 994–1008.

[7] J. Gani, "On a partial differential equation of epidemic theory," *Biometrika*, Vol. 52 (1965), pp. 617–622.

[8] ———, "On the general stochastic epidemic," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, 1967, Vol. 4, pp. 271–279.

[9] W. Gontcharoff, "Détermination des fonctions entières par interpolation," *Actualités Scientifiques et Industrielles*, Paris, Hermann, 1937.

[10] D. G. Kendall, "Deterministic and stochastic epidemics in closed populations," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1956, Vol. 4, pp. 149–165.

[11] W. O. Kermack and A. G. McKendrick, "Contributions to the mathematical theory of epidemics," *Proc. Roy. Soc. Ser. A*, Vol. 115 (1927), pp. 700–721.

[12] A. G. McKendrick, "Applications of mathematics to medical problems" *Proc. Edinburgh Math. Soc.*, Vol. 44 (1926), pp. 700–721.

[13] V. Siskind, "A solution of the general stochastic epidemic," *Biometrika*, Vol. 52 (1965), pp. 613–616.

[14] P. Whittle, "The outcome of a stochastic epidemic—a note on Bailey's paper," *Biometrika*, Vol. 42 (1955), pp. 116–122.